The versatility of segment routing in terms of deployment, distributed versus centralized, network types, data centers/WANs/access, and use cases makes it a solid option for end-to-end network deployment.

SEGMENT ROUTING

ACG RESEARCH PAPER

Ray Mota

TABLE OF CONTENTS

Note: V6 augments and expands the capabilities you had in the MPLS world

TABLE OF CONTENTS

Service providers and large enterprises face stiff challenges: the network infrastructure and their operations are growing at tremendous pace and becoming complex. IP/MPLS networks have become operation intensive because of their complex implementations. Service providers feel the added pressure of falling revenues and stiff competition by the over-the-top providers as well as the challenge to innovate. These drivers make the network owners think about a transport technology that can provide convergence across layers, domains and offload the complexities in the networks today.

One of these technologies is Segment Routing (SR), which has caught the attention of the network planners because of its potential to simplify and unify the transport layer. It is a source-based routing technology that enables IP/MPLS and IPV6 networks to run more simply and scale more easily. Segment Routing eliminates resource-heavy signaling protocols of MPLS and moves intelligence to the source/edge of the traffic thus removing complexity from the network. In the IPV6 networks, SR opens new possibilities of network programming and opens new avenues of flexibility, control and feature-rich use cases.

This paper details how each of these use cases can be implemented and describes the technology needed to understand the use cases[1].

Use Case	Technical Benefits	Business Benefits
Traffic Engineering (TE)	• Simple TE using stateless core, eliminating the need for complex RSVP-TE and complex TE configurations.	• Network and operational simplicity translates into lower capital expense (capex) and operation expense (opex). • Easy implementation of on-demand SR policy can unlock new business opportunity to sell service level agreements-based (SLA) services.
Protection Based on TI-LFA	• 100% coverage without micro loops against any failure (link, node, SRLG). Better than any IP protection today.	• Increase in network robustness and resilience. • Faster convergence time and increased network availability.
Network Programmability Using SR for IPV6 (SRV6)	• Instruction sets inside the SRV6 header enables network programming. • Convergence of services, overlay, and underlay into one IPV6 layer.	• Opens the network for innovation and new services beyond just connectivity. • Simple implementation of network function virtualization-based (NFV) service chaining.
SR Unified Fabric	• Uniform SR transport layer across access, metro, core and data center eliminates the need for complex reclassification at network boundaries.	• Opex and capex reduction by having uniform transport layer across access, metro, core and data centers.
5G Transport	• Potential to replace tunneling protocols such as GTP-U.	• Operational simplicity without the need for additional tunneling protocol. • Network slicing through inherent ability to TE and network programmability.
Software Defined Networking	• Flexibility to use SR in distributed, centralized and hybrid environments.	• Automatic traffic decisions can result in opex reduction. • Applications based traffic control on low latency, high bandwidth across access, core and data centers.

1 SR is no longer just in test labs; Microsoft, Vodaphone, Comcast, Walmart, China Unicom, Colt, and Bell Canada have implemented it.

Segment Routing is a promising technology that can be seamlessly deployed in today's MPLS and IPV6 networks. The versatility of the technology in terms of deployment (distributed versus centralized), network types (data centers or WAN), diverse use cases makes it a good candidate for deployment in any kind of WAN, data center, access, metro or virtualized environment.

THE FOLLOWING ARE RECOMMENDATIONS FOR NETWORK DESIGNERS, PLANNERS AND KEY DECISION MAKERS:

1   Assess the technological and operational pain points of current IP/MPLS networks and IPV6 networks. Recommendation: Bring on board the operation team in this exercise.

2   Understand the different use cases for SR. Every commercial deployment today has been driven by use cases. The biggest use case is simplicity and scalability.

3   For greenfield, it is easier and recommended to deploy SR because of the opportunities the technology offers, current IETF standards activities and success in real production networks.

4   For the brownfield environment, SR can be enabled in current IP/MPLS networks without any rip and replace strategy. It can co-exist with RSVP-TE/LDP.

5   Service providers can enable SR in their current networks on limited scale before global migration.

6   Implementing SR is a low-risk initiative considering that major protocols will be offloaded instead of burdening the network; ultimately the network will become simpler.

7   SR with a centralized controller makes sense as the core of the network is already stateless, and the controller can further take away the path computation burden off the edge nodes, enabling end-to-end control across multiple domains.

8   SRV6 enables flexible network programming. It enables the collapse of multiple layers and eliminates the need for overlay and additional protocols for service chaining, making the networks simpler to run and operate.

9   SR unified fabric leads to a simpler end-to-end transport network and reduces the number of transport protocols needed across access, metro, core and data centers.

10   SR is a way for the service providers to make their networks simpler and unlock new revenue potentials.

INTRODUCTION

Service providers' and large enterprises' networks are growing at a tremendous pace and becoming more complex and difficult to manage. This results in an increase in operation expenses (opex) and has forced the network owners to ask if there is a leaner and simpler way to manage and run networks. Is Segment Routing (SR) the answer to these network problems?

Segment Routing is a source-based routing technology for IP/MPLS and IPV6[2] networks. Although still in its infancy because the standardization activity is in progress, the industry is backing it up heavily. This can be gauged by the number of Internet drafts[3], which are under progress in the Internet Engineering Task Force (IETF).

This paper introduces Segment Routing concepts and their benefits. It describes several use cases to enable the business leaders/decision-makers in the networking industry to understand its emerging applications.

SEGMENT ROUTING DEFINED

Segment Routing is a source-based tunneling technology where a source chooses a path. The information is encapsulated in a packet header as an ordered list of segments, which sends the information, including the detour information, to its destination around the network.

SR can be applied in both IP/MPLS and IPV6 networks. In IP/MPLS networks it can be implemented without changing the data plane; in IPV6 networks it can be applied by adding a new routing extension header. SR when applied in IPV6 networks is also called SRV6.

INTEREST IN SEGMENT ROUTING

Vendors and service providers initiated the development of SR. In May 2016 the authors of RFC 7855[4] pointed out that the current networks could not easily fulfill requirements and there was a need to have simpler and flexible networks utilizing Segment Routing for:

- ✅ Network programmability
- ✅ Simplification and reduction of network signaling components
- ✅ Load balancing and traffic engineering

---

2 Illustrations are presented with IP/MPLS, but they equally apply to IPV6.
3 The are many active Internet drafts on which the IETF is currently working. The architecture and use cases are handled in the SPRING working group, which is dedicated to Segment Routing. Protocol extensions are handled in their respective working group: ISIS, OSPF, IDR, PCEP, and 6MAN.
4 Source Packet Routing in Networking.

BENEFITS OF SEGMENT ROUTING

Simplicity
Simple to operate, maintain and troubleshoot

Scalability
Scalable as the network core does not keep any state information allowing the core to scale

Fast reroute
Guaranteed 50 m sec. Protection in all cases: link, node, srlg

Segment Rounting

Traffic engineering
Complete control over how the traffic is routed in distributed or centralized control environment

Network programmability
SR for IPv6 (SRv6) takes segment routing to the next level by bringing network programmability
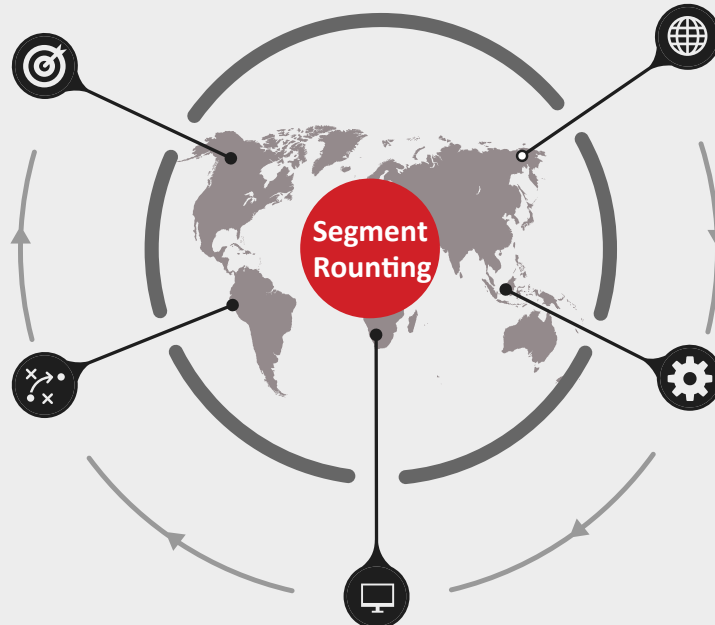
Figure 1. Benefits of Segment Routing

SIMPLE, SCALABLE AND ECMP FRIENDLY

LDP and RSVP-TE are the de-facto signaling and label distribution protocols in IP/MPLS network used for years, but are they scalable?

SCALABILITY ISSUES OF LDP AND RSVP-TE[5]

Control Plane Sessions: For LDP each router maintains sessions (LSPs state), which are equal to the number of LDP neighbors. For RSVP-TE, the number of sessions is equal to the total number of LSPs in which the router is involved (whether ingress, egress or transit). In the RSVP-TE case, if a topology includes N fully meshed routers, there will be a need to maintain a state of N x N (N square) LSPs in each router. This quickly runs into an N square problem because the number of N increases. From a control session perspective, RSVP-TE can run into scalability issues.

5 https://books.google.com.sa/books/about/MPLS_Enabled_Applications.html?id=2lxbaQ-VN8sC&redir_esc=y.

FORWARDING STATE

LDPs maintain forwarding state of all Forwarding Equivalence Class (FEC) in the network, because each FEC is reachable by any other LDP router in a network. RSVP-TE only keeps the forwarding state of the LSPs that traverse through it and potentially their protection path. From forwarding state perspective, LDP runs into scalability issues if a network becomes extremely large.

RSVP-TE can also perform traffic engineering in IP/MPLS networks; however, it involves complex tunnel configurations on interfaces and is difficult to troubleshoot. LDP cannot do traffic engineering, but it can lose synchronization of the LDP and IGP because LDP depends on IGP for route convergence.

SR is scalable[6] because it does not rely on LDP/RSVP-TE, and there is no need of keeping thousands of labels in an LDP database. It avoids thousands of MPLS traffic engineering LSPs in the network.

SR uses extensions to existing IGP protocols for signaling purpose. Relying on IGP has other benefits too; it can take advantage of Equal Cost Multi-Path Routing (ECMP) to load balance across multiple available paths in the network and gain better bandwidth utilization. This kind of flexibility does not exist in current RSVP-TE, which would need complex manual configurations for ECMP functionality.

THE FOLLOWING TABLE SUMMARIZES THE KEY BENEFITS OF SEGMENT ROUTING VERSUS MPLS WITH LDP/RSVP-TE:

	Use Case	Technical Benefits	Business Benefits
	Operational Simplicity	• Very simple	• LDP is simple but RSVP-TE is complex
	Scalability for TE	• High-As minimal status in the network. State is in the source node	• Low-RSVP needs to create full mesh LSPs and middle nodes need to keep a lot of transit information
	Fast Reroute	• Yes (100%)	• Yes (Close to but not 100%)
	Number of Protocols	• Signaling protocols like LDP or RSVP-TE are not needed. Less protocols in network	• LDP or RSVP-TE or both are always needed
	ECMP for TE	• Yes	• No
	Traffic Engineering	• Yes and Simple	• Only RSVP-TE, but more complex
	SDN	• Supported	• Only RSVP-TE

Table 1. MPLS with Segment Routing versus MPLS with LDP/RSVP-TE

---

6 https://www.cisco.com/c/en/us/td/docs/ios-xml/ios/seg_routing/configuration/xe-3s/segrt-xe-3s-book/intro-seg-routing.pdf.

TECHNOLOGY BEHIND SEGMENT ROUTING

SR uses segments and segment identifiers (SID). A segment is a basic unit in SR. By combining multiple segments, an end-to-end route can be created. If traffic needs to go from ingress at A to egress at H with a diversion at E, then the three segments are enough to define the path (Figure 2). Additionally, there be should some identifier associated with the segment; this identifier is called Segment Identifier. The end-to-end path is also sometimes denoted as a SID list (SID 1, SID 2, and SID 3).
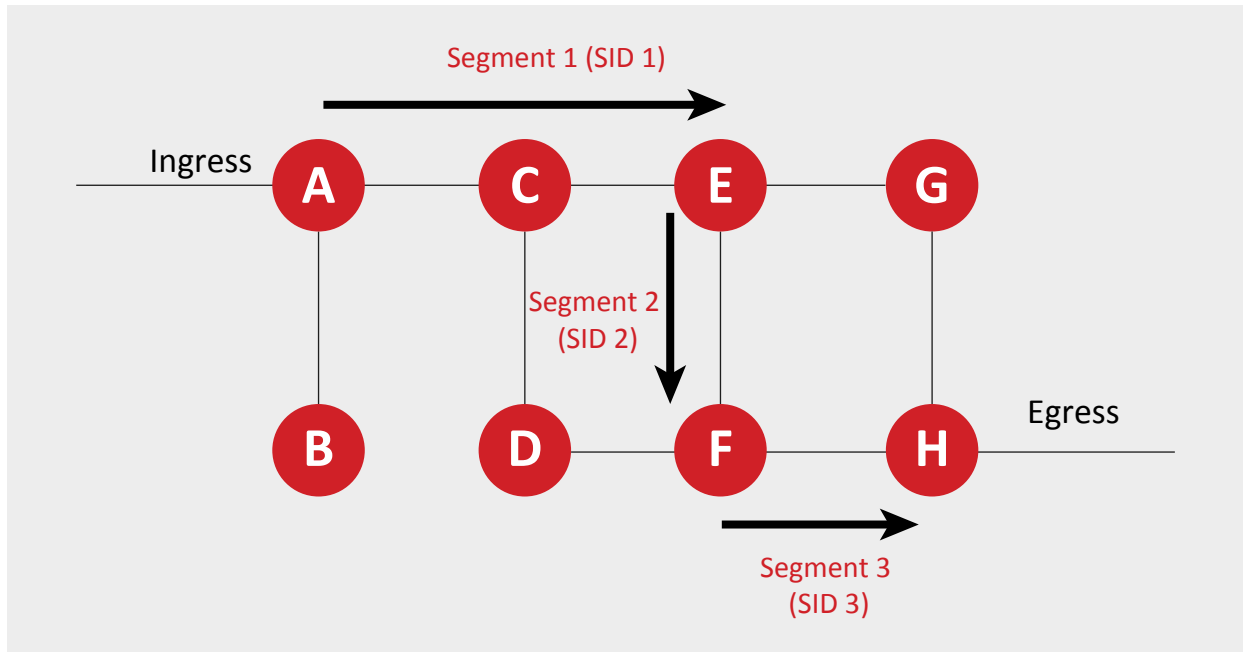
Figure 2. Segments in Segment Routing

In MPLS, a segment is encoded as an MPLS label. A stack of labels represents an ordered list of segments. The top label is the one that is processed by the node that receives it. Upon processing the packet, the top label is popped from the stack.

In IPV6 a new routing header is defined to enable Segment Routing. A segment is encoded as an IPV6 address. An ordered list of IPv6 addresses represents an ordered list of segments. The destination address of the packet shows the active segment.
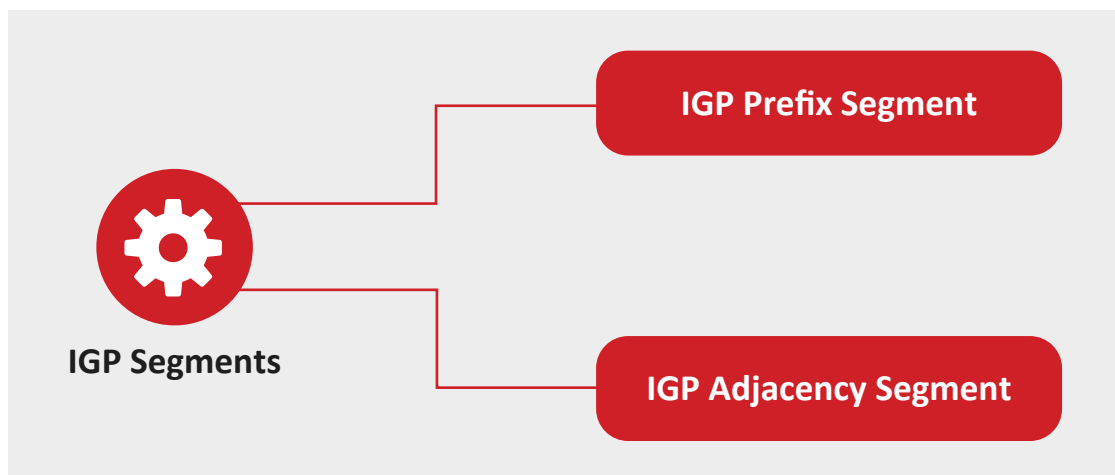
TYPES OF SEGMENTS

Figure 3 - Types of IGP Segments

Within an SR domain, an IGP node advertises segments for its attached prefixes and adjacencies. These are called IGP segments. Advertisements of IGP segments require extensions to link-state IGP protocols such as OSPF and IS-IS.

IGP PREFIX SEGMENT, PREFIX SID:

- IGP Prefix Segment depicts a path to an IGP Prefix. It is an ECMP aware segment. Its segment identifier is called Prefix SID.

- The SID value (which is unique within SR domain) is allocated from a unique pool called the SR Global block (SRGB).

- Every router is identifiable by a unique Prefix SID in the network so that other routers know where to send the traffic once it sees that SID.

- When SR is used in MPLS, Prefix SID is allocated in the form of an MPLS label. When SR is used in IPv6, Prefix SID is allocated as an IPv6 address.

IGP-Node Segment, Node-SID: IGP-Node segment is a special subtype of Prefix Segment. The Node Segment signifies a path to a node (for example, a loopback) in an IGP domain; it is identified by Node SID value, which is unique in the SR domain.

IGP-Anycast Segment, Anycast SID: IGP-Anycast Segment is a special type of Prefix Segment that shows ECMP aware path toward the closest node of anycast set. It points to a group of routers with a common SID value called Anycast SID.

IGP-Adjacency Segments: These are local to each node and are installed and advertised only on directly connected neighbors, identifying a specific adjacent link. IGP-Adjacency Segment is identified by Adj-SID, which is dynamically allocated by a node (outside the SRGB block). If a router has four adjacent links, it will allocate a unique Adj-SID to each one of them. Once it sees that Adj-SID in the incoming label stack, it knows on which link the traffic should be forwarded.

Figure 4 **Prefix and Adjacency Segments in IP/MPLS Segments**
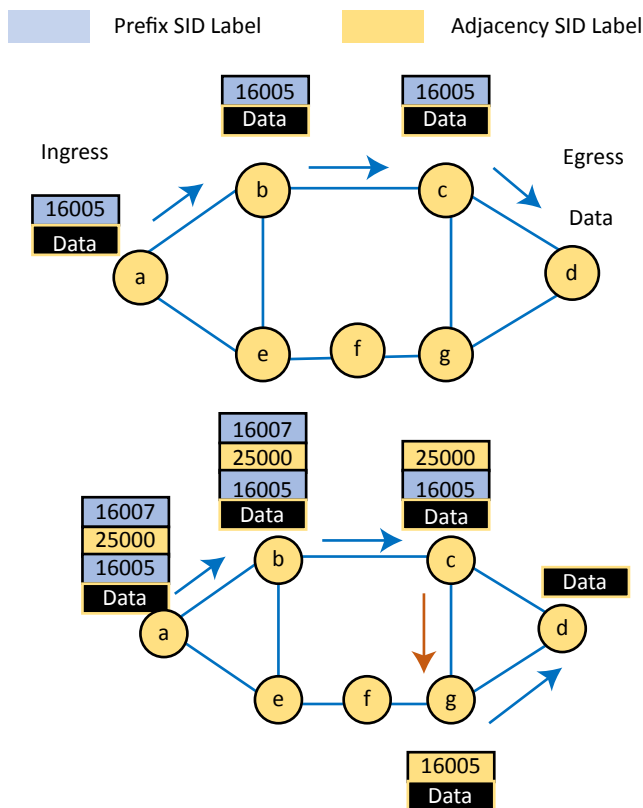
IGP PREFIX SEGMENT

- 16005 is the prefix SID for router d.

- Label '16005' is added at ingress and every node in the path knows that packet needs to be forwarded to router 'd'. Label is swapped.

- Label is swapped to same value at each node giving impression that label is untouched.

- At router'c' label 16005 is popped and packets sent to d.

IGP ADJACENCY SEGMENT

- Implementing same route as above but now inserting Adjacency segment at node'c', to see how it can change the route at 'c'.

- 16005, 16007 are prefix SIDs for 'd' and 'c' respectively.While 25000 is Adj. SID for section between 'c' and 'g'.

- 'a' outputs 16007 as top label so that traffic can be sent to 'c', 'b' just swaps the top label to same value and sends to 'c'. At c it is popped and the next label is 25000 which is further popped as it is the Adj. SID, pointing to neighbor 'g'.

- Router 'g' sees Prefix SID 16005 of router 'd', it is popped and traffic sent to router 'd'.

BGP SEGMENTS

There are two kinds of BGP segments allocated and distributed by BGP.

BGP Segments

BGP Prefix Segments
BGP Prefix ID

BGP Prefix Segment depicts a route to a BGP Prefix. A BGP Prefix SID identifies the BGP Prefix Segment. It is unique within an SR domain. BGP has been extended to carry segment routing SID. It is an ECMP aware segment.

BGP Peering Segments
BGP Peering SID

BGP Peer Segment helps in identifying a particular BGP peer link among several available peer links. This greatly helps in BGP Egress Peer Engineering (EPE). An EPE enabled egress router may advertise segments corresponding to its attached peers. These segments are called BGP Peering Segments (with the ID as BGP Peering SID).

Figure 5. BGP Segments

This helps in traffic engineering and routing toward the desired BGP peer. A controller must have visibility to BGP Peering Segments and external topology of egress border router. BGP-LS is used for signaling BGP Peer SIDs to the controller. It has local significance and is dynamically allocated by the signaling router.

SR Controller

a DC (BGP SR)   WAN (IGP SR)   25000   PEER

16001   16002

b   c   AS2

AS1   AS3

Figure 6. SR Controller

BGP prefix and peering Segments
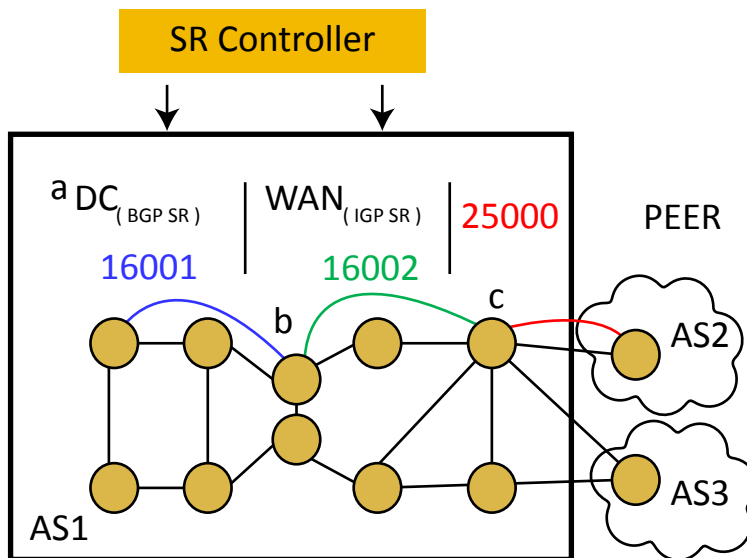
- There is a requirement to engineer path for a particular low latency application from data center to cross WAN and then exit towards a preferred path which is through AS2
- To reach from Data Center ('a') to particular egress peer (AS2) will need threee segments which are stacked as labels. {16001, 16002 and 25000}
- 16001-> BGP Prefix SID to reach from 'a' to 'b'
- 16002-> IGP Prefix SID to reach from router 'b' to egress router 'c'
- 25000-> BGP Peer SID to select particular link that is directly connecting 'c' to AS2

SEGMENT ROUTING USE CASES

Use Case: Fast Re-Route (Topology Independent LFA)

SR runs in service provider networks that provide mission-critical services, which require recovery from failure that is quick, simple and predictable. It is a default requirement to have failure recovery in less than 50 milliseconds.

There has been continuous improvement in the resilience mechanisms in IP/MPLS networks: RSVP-TE-Fast reroute, Loop Free Alternate (LFA) and remote LFA, which has seen wide adoption. Although mechanisms have improved, there are none that can guarantee 100% coverage for all failure scenarios. It is not uncommon to see that LFA converges on a path that is suboptimal. SR solves the issue of micro-loops[7] that may happen in in LFA. SR utilizes Topology Independent LFA (TI-LFA)[8] that can provide loop-free guaranteed coverage against link, node and local SRLG failure in 100% of cases.

Advantages of TI-LFA

100% Coverage
100% survival in link, node and SRLG Failure

Simple & Scalable
No need for targeted LDP sessions. No state in network

Zero Micro loops
No issues of micro Loops in transient conditions

TI-LFA BENEFITS

Optimal Post Convergence Path
Protection path will automatically follow the natural IGP backup path instead of sub-optimal path
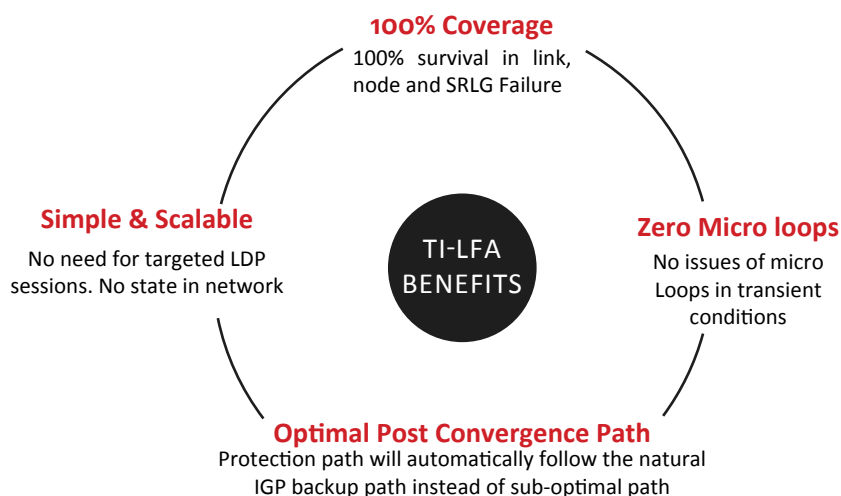
Figure 7. Advantages of TI-LFA

---

[7] First time in 50 years of IP existence that a solution to this root IP problem of microloops7 is proposed using SR.
[8] Topology independent refers to the ability to provide a loop-free backup path irrespective of the topology before and after the failure.

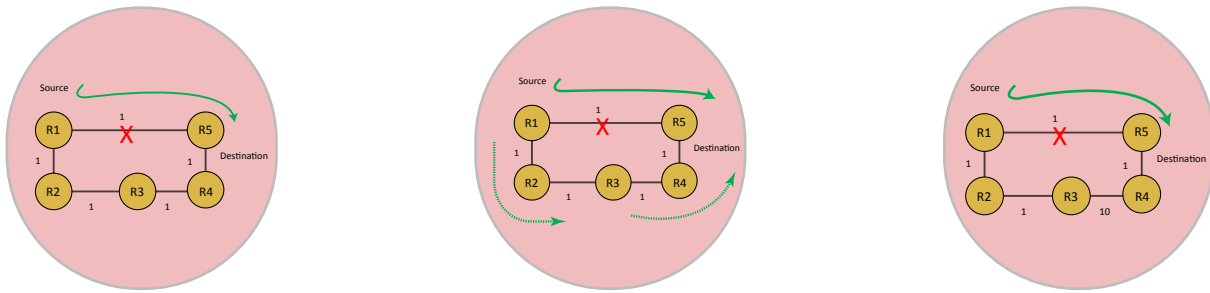## TO APPRECIATE THE ADVANTAGES OF TI-LFA, IT IS IMPORTANT TO UNDERSTAND THE SHORTCOMINGS OF LFA AND REMOTE LFA.

Figure 8. Issues with LFA

WHAT LFA CANNOT SOLVE AND ISSUE OF MICRO-LOOPS !	HOW REMOTE LFA SOLVES LFA ISSUES !	WHAT REMOTE LFA CANNOT SOLVE !
• LFA has issue when used in ring with more than 3 nodes.	• Remote LFA chooses a next hop router to tunnel protecting traffic, that will not send it back to itself.	• In this scenario, IGP metric between R3 and R4 is increased to 10. (also called double segment example)
• When primary link between R1 and R5 fails, traffic is immediately diverted to R2.	• That is a router two hops away i.e. R3 (Also called PQ Node as per RFC 7490)	• There is no PQ node. If R1 sends to traffic to R3 through targeted LDP, it will send it back to R1 as the shortest metric path from R3 to R5 is through R1
	• For backup path,R1 will create targeted LDP session to R3 so traffic can get through R2.	
• However R2 will send the traffic back to R1 as shortest metric path of R2 to R5 is through R1. this will create micro-loops until IGP converges	• Once traffic reaches R3, it can easily go to R5 as that in shortest metric path.	• Remote LFA is not able to solve scenarios like these ones

LFA does not provide 100% coverage; however, Transport Independent LFA with Segment Routing does not need any targeted LDP session. This makes the protocol very simple and scalable.

In the same scenario in Case 3, for the protection route, at the ingress router R1, three segments are built using three SID labels:

01 Prefix R3 label to send the traffic to R3

02 Adj. R3-R5 label to send traffic from R3 to R4

03 Prefix R5 label to send traffic from R4 to destination R5

Using Adjacency SID at Node R3 has solved the issue of crossing the high metric link from R3 to R4; when the traffic reaches Router R3 and sees the Adjacency SID label Adj R3-R4 it immediately knows that it needs to send the traffic on the adjacent link to R4 irrespective of the metric on this link. Segment Routing has solved the protection problem easily by building three label stacks without any need of targeted LDP session.
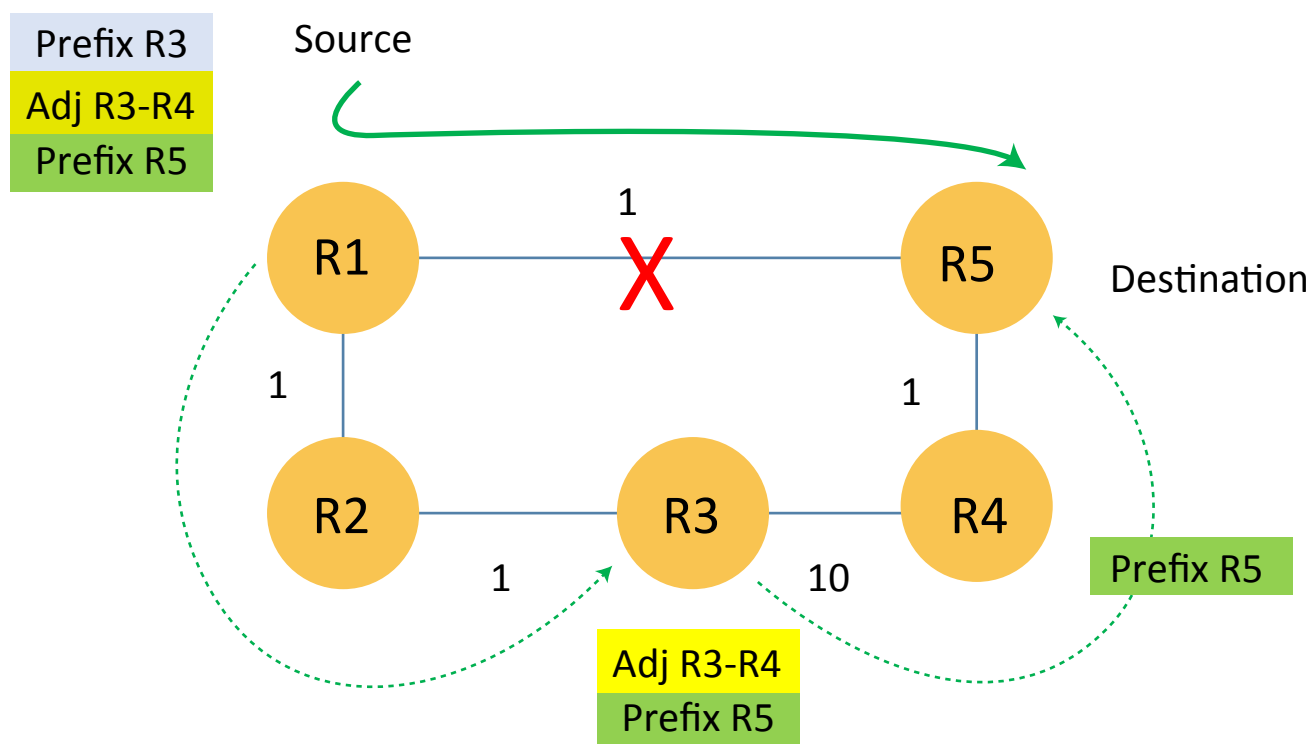
Prefix R3
Adj R3-R4
Prefix R5

Source

1

R1

X

R5

Destination

1

1

R2

R3

R4

1

10

Adj R3-R4
Prefix R5

Prefix R5

Figure 9. TI-LFA Solution

In addition to link failures, Segment Routing also provides node and SRLG protections9.

NODE PROTECTION

A user has the option to enable node protection using TI-LFA; if enabled, the post convergence backup path does not consider the next hop neighbor while calculating the backup path.

SRLG PROTECTION

Shared Risk Ling Group (SRLG) is a situation in which links share a common physical infrastructure (for example, using common fiber cable). These links carry shared risks. Once a link breaks, it is expected that traffic is converged on a backup path that excludes the shared risk group for the protected link.

TI-LFA SRLG protection functionality finds a backup path that excludes the SRLG of the protected link; consequently, there is no risk of SRLG failure. TI-LFA can solve all protection issues in less than 50m seconds without any risks of micro-loops.

Remote LFA sometimes does not converge on an optimal path on the protection route. In the scenario in Figure 10 there is a break on the primary link between Router R1 and Router R5 (green path). The remote LFA chooses R3 (PQ node) as its next hop for protection traffic (red path) although its metric is high instead of sending the traffic through R2, which is low metric link. It cannot send it to R2 because the shortest path of R2 to the destination is through R1, which would create micro loops. TI-LFA solves this issue by stacking two Prefix SIDs (blue path). The top label points to R2, forcing the packet to choose R2 as its next hop and the next label (Prefix SID) points to destination R5, easily solving the suboptimal convergence issue. TI-LFA follows the natural path of the IGP convergence. In the example, traffic from R1 to R5, if R1-R5 link is not available, is always through R2.
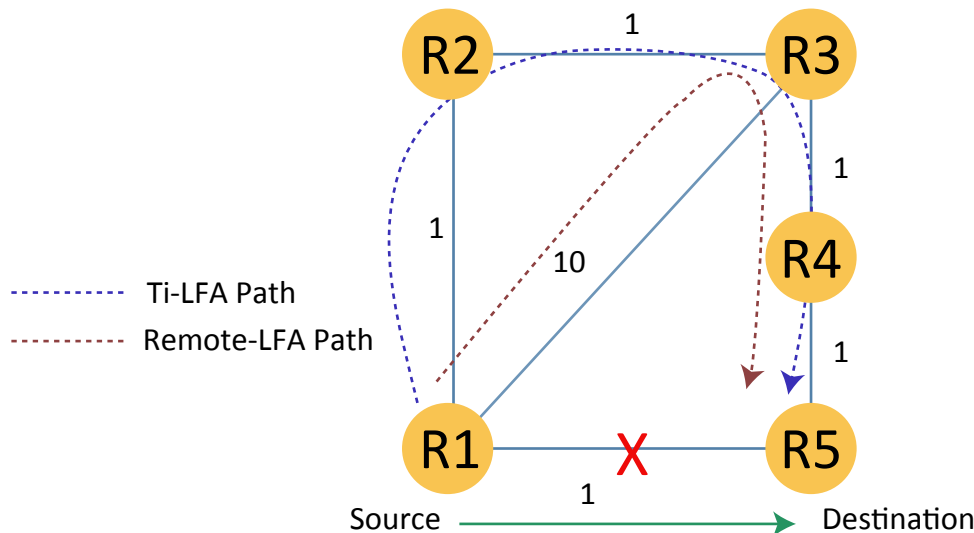

Figure 10. Suboptimal Path Convergencet

USE CASE: TRAFFIC ENGINEERING (SRTE)

SRTE has become one of the widely adopted use cases for Segment Routing because of the simplicity and scalability it provides. Traffic engineering in MPLS until now has rarely been implemented in large service provider networks because of the complexity. Not only does SRTE provides simplicity and scalability but also provides an SR native way of implementing traffic-engineered paths that take advantage of the ECMP behavior of IP. Complexity is further reduced in the network because of the additional benefits of automation through on-demand SR policy implementation and automated traffic steering in the network.

To understand the benefits of SRTE it is important to understand the concept of SR Policy and Binding Segment.

SIMPLE & SCALABLE
No state network. No need for complex tunnel configurations. policy is in the SR header

MULTI DOMAIN CAPABLE
Works across multiple domains to implement end to end traffic engineered paths

SRTE Benefits

SR NATIVE
IP Optimized, less number of SID labels, ECMP native

ON DEMAND SR POLICY & AUTOMATIC TRAFFIC STEERING
Flexible and On demand SR policy implementatiopn without pre configuration and automatic traffic steering as per policy

Figure 11. Advantages of SRTE

SR Policy and Binding Segment (SID)

An SR policy is identified by a tuple:

- Headend, where policy is initiated

- Endpoint, which is the destination of the policy

- Color, an arbitrary numerical value that shows different policy types, for example, green for low-latency path; red for high bandwidth path

In the following case two different policies are configured: green with low latency and red with higher bandwidth. A policy can have multiple candidate paths. For example, red policy has two paths. The preferred candidate path among multiple candidate paths is identified by the highest preference number (one of the parameters of candidate paths) among them.
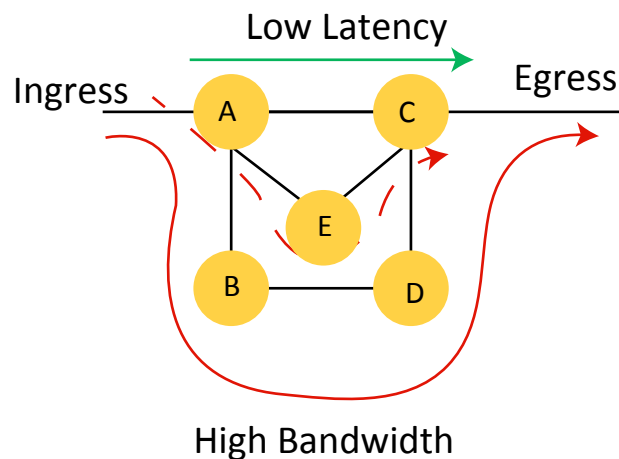

Figure 12. SR Policy in Segment Routing

Binding Segment (SID)

Binding Segment is a new type of segment for traffic engineering, also identified as BSID, which is fundamental for SRTE and brings scalability and service independence to Segment Routing.

A candidate path in SR is identified by its BSID number. When an ingress router receives labels with top of the stack as BSID, it will pop the top label and push the policy. In the example, BSID has a value of X with the following candidate path for red: BSID list = X= {Prefix B, Prefix D}.

If Router A receives a packet with the label stack as {X, Prefix C}, it knows that X is BSID. It will pop this label and add the candidate path to the label stack. This will result in the label stack (Prefix B, Prefix D, Prefix C), which follows the red route.

SRTE Advantage: Simple and Scalable

RSVP-TE, the protocol for traffic engineering in IP/MPLS, is not popular because of the need to create a lot of tunnel configurations for TE policies. It quickly runs into scalability problems because of these issues.

SRTE keeps core very light and scalable (as core is stateless). SRTE supports both explicit routing and constraints-based routing such as RSVP-TE. Using constraints-based routing, flexible policies can be created automatically in centralized and distributed environments based on latency, disjoints and preferred paths, etc.

SRTE Advantage: SR Native Algorithm

Instead of using algorithms of RSVP-TE to calculate the best path based on constraints, SRTE uses SR Native Algorithm. This results in label stack reduction and a path that is load balanced because SR has ECMP native capabilities.

In Figure 13, conventional TE mechanisms are compared with SR Native TE mechanisms to find a traffic engineered path between A and F that does not pass through the red link. In conventional TE mechanisms the single best path is calculated through B, C, D, and E routers.

This is more like a classical TDM approach in which one path (instead of multiple paths) is utilized for traffic transfer and does not take advantage of load balancing across equal cost links (to have ECMP in RSVP-TE would require additional configurations and complexities).

Because SR natively supports ECMP, it utilizes all equal cost paths, which also results in fewer segments. For example, just pointing to Prefix E, will load balance the traffic across available three links. By using just two SIDs E and F as label stacks, traffic can reach destination F, utilizing three paths.

SR achieves the traffic engineering objectives with an SR native approach and with fewer label stacks (shorter SID list).
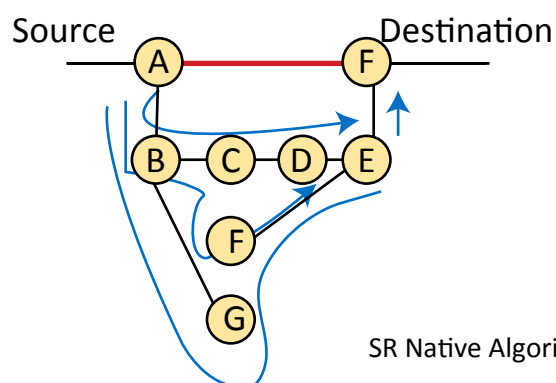

Source   Destination
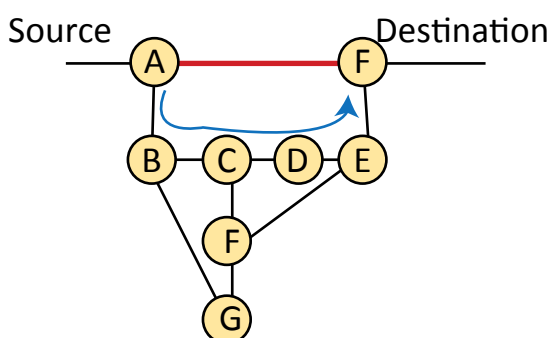
Classis RSVP-TE Algorithm
Single path
No ECMP
Long SID list {B,C,D,E,F}

SR Native Algorithm

Load balancing with ECMP

Short SID list {E,F}

Figure 13. RVSP-TE Algorithm versus SR Native Algorithm

SRTE has a novel way of instantiating SR policy on demand instead of configuring it beforehand. SR policy can be instantiated on demand based on BGP Next Hop. This creates a very dynamic, flexible and automatic way to apply policies. Not only can the policy be instantiated on demand, but traffic can be automatically steered (because of BSID) based on the forwarding plane set by the on-demand SR policy.

The on-demand policy makes traffic engineering very simple, automatic and lightweight. This contrasts with RSVP-TE that needs to have policies preconfigured with complex tunnel configurations. For example, on-demand SR concept with automatic steering: Customer A buys a premium low-latency service and Customer B a basic VPN service (lowest IGP metric) from the service provider.

The challenge is to configure policies on the fly once BGP routes are installed and then steer traffic according to the new policies. Router R5 at the destination advertises two BGP routes: green for low latency VPN and red for basic VPN based on lowest IGP metric. Once the SRTE process at R1 sees two different colors in the BGP advertisement, it will create two policies:

Green color -> SID List for low latency path = {Prefix R4, Adj. R4-R5} with BSID 1000

Red color -> SID List for Low IGP metric path= {Prefix R5} with BSID 2000

These two policies/SID list are created on demand once Router R1 receives BGP advertisements from R5 with the colors and then placed in the forwarding table of the router R1.

The second process is the automatic steering of the traffic according to these policies without any prior configuration. Once each customer starts sending traffic to its remote site, the VPN traffic path will be resolved according to the relevant BSID, automatically steering the traffic to desired path of the VPN.

This kind of automation enables SRTE to scale on demand to steer traffic according to the customer's SLA or requirements of the applications.
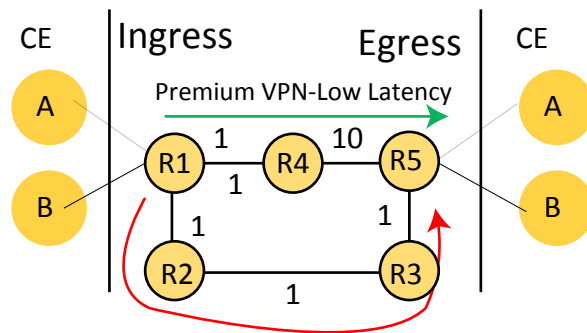
Basic VPN-Lowest IGP metric

Figure 14. On-Demand SR Policy and Automatic Traffic Steering

SRTE Advantage: Multidomain Capable

SRTE is multidomain capable and designed in way that it can run in a multidomain environment with or without a centralized controller. To validate paths and compute dynamic paths, the SRTE process maintains an SRTE-DB that can run flexibly in a headend router or a centralized controller. The attached domain topology can be learned via IGP, BGP-LS or NETCONF. A nonattached (remote) domain topology can be learned via BGP-LS or NETCONF. In a centralized environment, automated PCE assistance can create end-to-end uniform policy-based constraints such as latency, disjoints and SRLGs.
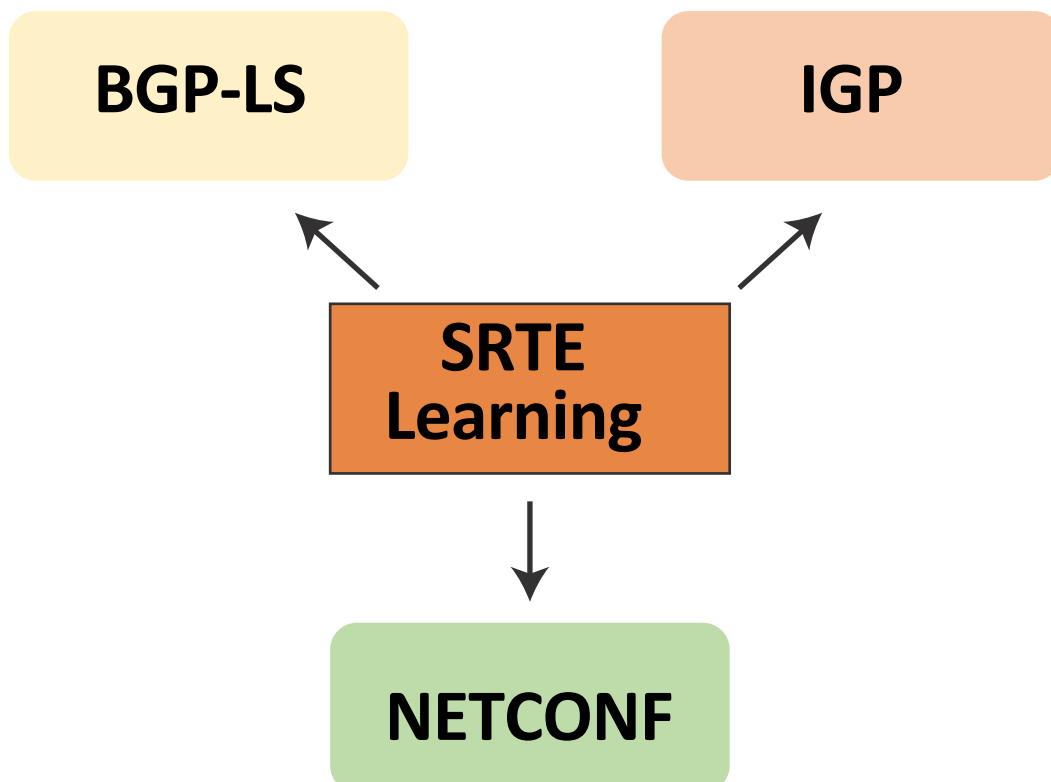

Figure 15. SRTE Learning

Use Case: Flex Algorithm

Flex Algorithm adds to the capabilities of SRTE. It does so through a new prefix segment, which is defined for a common objective such as minimize the igp-metric/delay or, for example, avoid a certain SRLG.

Many different types of constraints can be defined. For example,, in a network with dual plane, a constraint would be to use a certain plane and avoid the other plane. SR allows computing paths with these constraints using certain algorithms. It then allows Prefix SID to be associated with these algorithms. They are called Flex Algorithms.

To provide maximum flexibility there is no strict mapping between the set of constraints and the algorithm associated with it. The mapping between the algorithm value and its meaning is flexible and defined by the user. The only requirement is that the routers participating in the domain should have common understanding of the algorithm value, hence the name flexible algorithm.

Flex Algorithm provides a new prefix segment to achieve one of the following objectives:10

Minimize the igp-metric or delay or TE metric
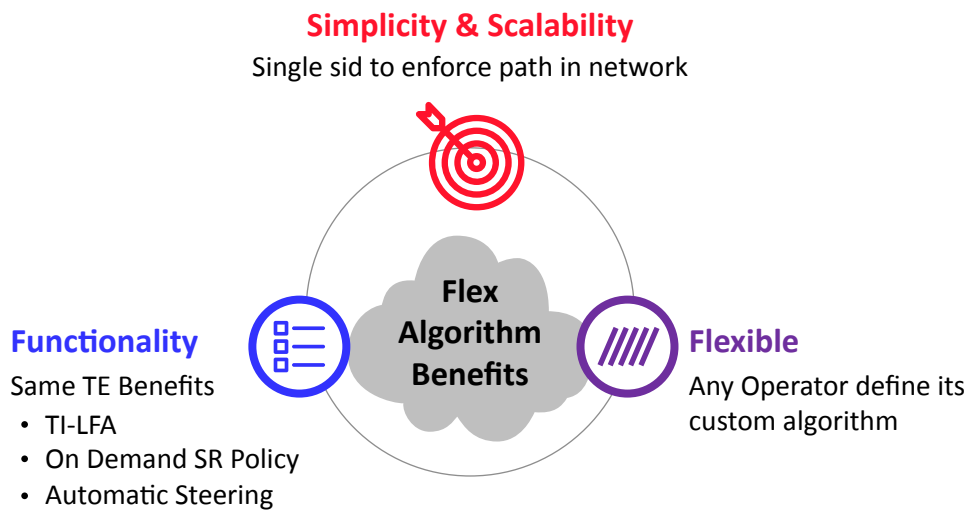
Avoid SRLG or affinity

Simplicity & Scalability

Single sid to enforce path in network

Flex Algorithm Benefits

Functionality

Same TE Benefits
- TI-LFA
- On Demand SR Policy
- Automatic Steering

Flexible

Any Operator define its custom algorithm

Figure 16: Benefits of Flex Algorithm

One of the interesting use cases for Flex Algorithm is the implementation of dual plane connectivity. Traffic is divided between two different planes. In the following the Flex Algorithm128 is associated with the red plane and Flex Algorithm129 is associated with the green plane.

The definition of these algorithms can be something like the following:
- Flex Algorithm128 = Minimize IGP metric and avoid TE affinity green
- Flex Algorithm129 = Minimize IGP metric and avoid TE affinity red

Routes with Prefix SID 128 will stay in the red plane and the routes with Prefix SID 129 will stay in the green plane. Even in case of fiber cut, for example, the protection route for red plane will stay in the red plane and the same for the green route. It is clear that the use case of dual plane connectivity can be achieved very easily by just using one Prefix SID defined through flexible algorithm.
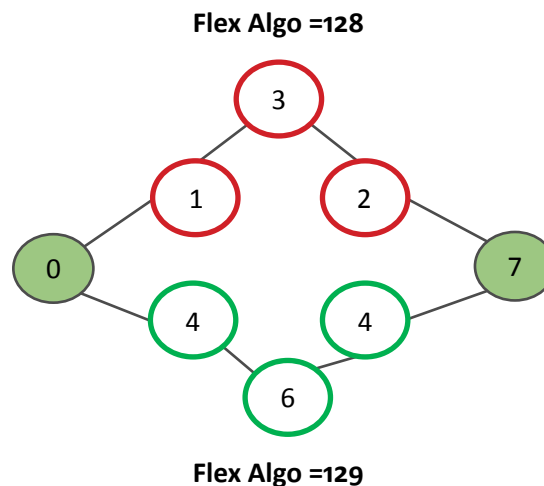
Flex Algo =128

Flex Algo =129

Figure 17: Dual Plane Connectivity Using Flex Algorithm

USE CASE:
SOFTWARE DEFINED NETWORKING

SR can be used in a centralized, distributed or hybrid environment. In a distributed scenario, the segments are allocated and signaled by IS-IS or OSPF or BGP. In a centralized scenario, the segments are allocated and instantiated

11 https://tools.ietf.org/pdf/draft-ietf-spring-segment-routing-13.pdf.

www.acgcc.com

by an SR controller (SDN controller). Although distributed and centralized intelligence can be combined for a hybrid scenario, for example, distributed can be used in the same IGP domain. When the destination is outside the IGP domain, the SR controller (SDN controller) may compute a source-routed policy on behalf of an IGP node.

When used in centralized environment, Segment Routing gives a de facto SDN architecture as there is no state in the network (only on the edge), and the traffic paths are computed and programmed by a centralized controller, which is usually Path Computation Engine.

By logically centralizing the control of the network, it is possible to program per-flow routing based on TE goals. With limited state in the network, SDN centralized controller can actively collect topology information from the network using existing protocols such as BGP-LS and then compute the best paths based on the constraints defined by a user.

Not all applications have equal value. Some are delay sensitive (financial transactions and VoIP); some are bandwidth intensive (data centers replication); and others need low jitter (video). Rather than manually configuring these tunnels, which may run into thousands, and managing them, such tasks can be handed over to a centralized SDN controller[12] . By integrating them with the application layer that can tell the requirements to the controller about SLA needs for the end-user applications, the controller can react in an agile way to the application routing in the network.
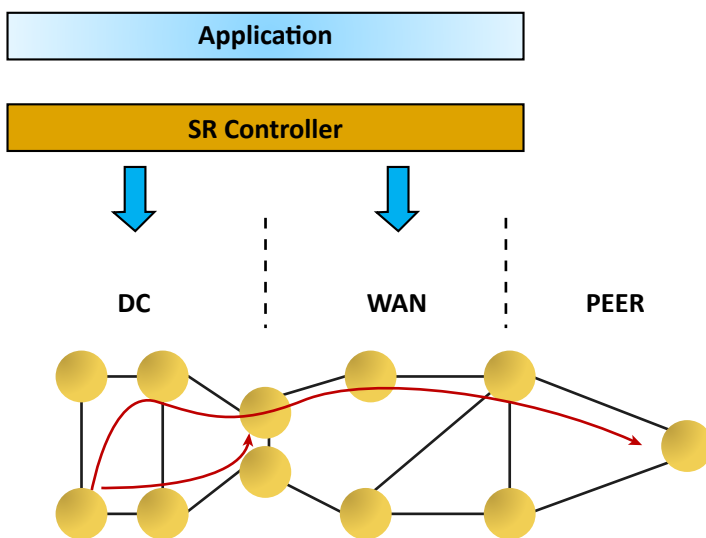
Using an SDN controller with SR expands traffic engineering possibilities, for example, setting up end-to-end policies across independent data center metro, access and backbone domains. It allows for complex protocol conversion between network domains and brings high scalability in the network.

Figure 18. SDN Controller in SR Environment

SRV6 USE CASE: NETWORK PROGRAMMABILITY

SR in IPV6 (SRV6) opens new paradigms that go beyond simple networking expected from SR. It brings the concept of instruction sets (functions) that enables complex network programming models.

Network programming can result in the collapse of technologies. One example is service chaining in NFV, in which a packet must travel to different service nodes (virtual machines) and perform different functions. In the presence of SRV6 underlay[13] and the capabilities of SRV6 to support instructions in SRV6 header, overlay and service layer functionality can be implemented easily in SRV6. This means there is no need for additional layers as SRV6 can eliminate both the overlay and the service layer and replace it with additional SRV6 headers to perform the functions of these layers. The network becomes simpler and run with fewer protocols.

---

12 https://www.nil.com/en/networking/segment-routing/s.
13 http://www.segment-routing.net/conferences/2017-nanog-network-as-a-computer-srv6/.

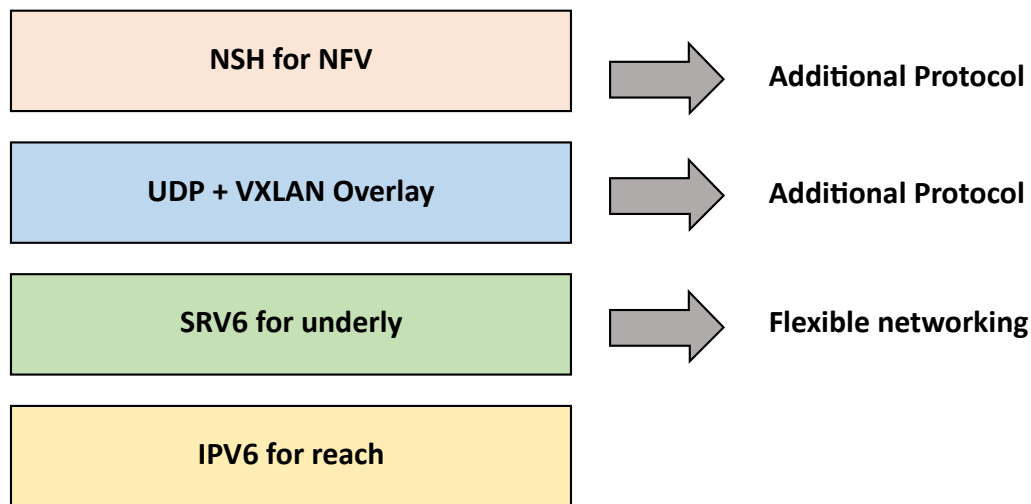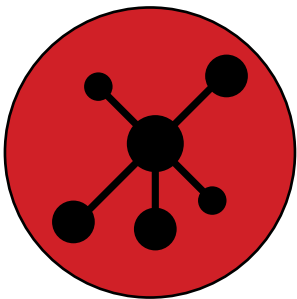| NSH for NFV | → | **Additional Protocol** |
UDP + VXLAN Overlay	→	**Additional Protocol**
SRV6 for underly	→	**Flexible networking**
IPV6 for reach		

Figure 19. SRV6 Replacement for Additional Protocols

HOW SRV6 ACHIEVES NETWORK PROGRAMMING?

Network programming is one of the powerful features of SRV6 and enables different functions to be associated with the SIDs in SRV6[14]. In Figure 20 the SRV6 SID has two parts: a locator part that identifies the address and a corresponding function part that is an instruction executed at the location described by the locator part. There is also a metadata TLV attached as a global argument to the SRH header that can be used to carry additional information, for example, credentials and performance information.

A router inspects the segment header only when a packet is addressed to itself based on the destination address of the packet. Functions with locators are stacked one above the other at the source node and sent to the network. After the first function is executed, the packet is sent to the second locator to execute the second function.
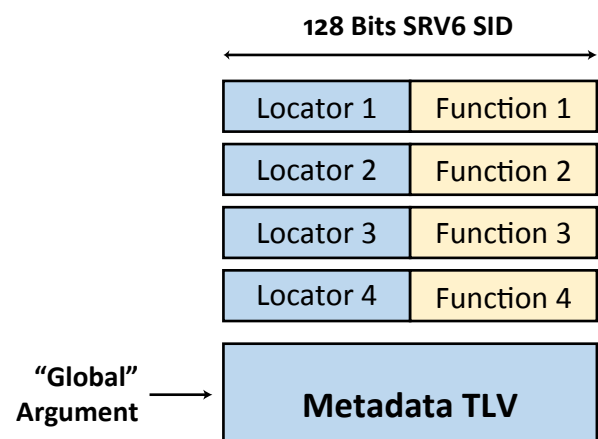
128 Bits SRV6 SID

Locator 1	Function 1
Locator 2	Function 2
Locator 3	Function 3
Locator 4	Function 4

"Global" Argument →

Metadata TLV

Figure 20. Global Argument

The behaviors of these functions are entirely up to the implementer. For example, these functions can be forwarding, encapsulation, decapsulation, L2 or IPV4 cross-connect, instantiation of SRV6 policy or combination of these actions. Any function can be attached to the SID. These stacks of segments act like a network program that can treat packets in different ways, going beyond just the forwarding functions that normal transport does. Very complex network functions can be executed in the network through the network programmability features of SRV6.

---

[14] https://tools.ietf.org/html/draft-filsfils-spring-srv6-network-programming-00#section-4.

SRV6 USE CASE: STATELESS SERVICE CHAINING

Service Function Chaining (SFC) is the process of steering traffic through an ordered list of functions, for example, load balancer, firewall and proxy.

SFC is defined by IETF. Network Service Header (NSH)[15] is part of SFC. It is imposed on packets or frames to realize service function paths. SR can achieve SFC because it can execute one function after the other based on the SRV6 header stack. However, since SR is inherently stateless and policy is only encoded at the ingress of the network, it is more scalable compared to NSH, which relies on the state configured at every hop of the network.

In Figure 21, the service chain shows different functions running on either virtual machines or containers in an NFV environment. Two different service chains are created. The applications can be SR aware or not (SR case proxy function can be used).
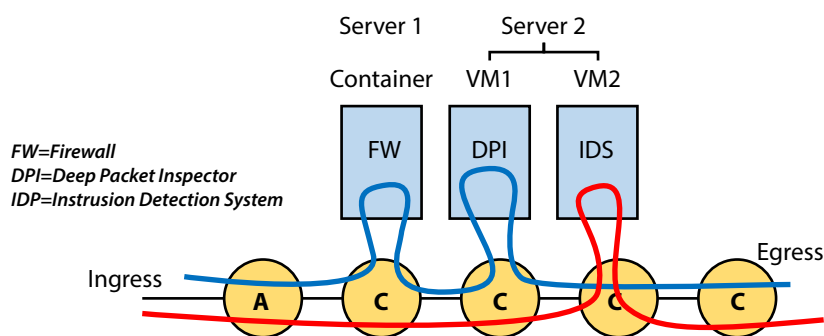

Figure 21. SRV6 Service Chaining

SRV6 is flexible and achieves service chaining by just programming the network header. SRV6 header is programmed at the ingress node with stackable instructions to be executed at each location in the service chain thereby eliminating the need for any state in the network. Consequently, it scales much better than NSH.
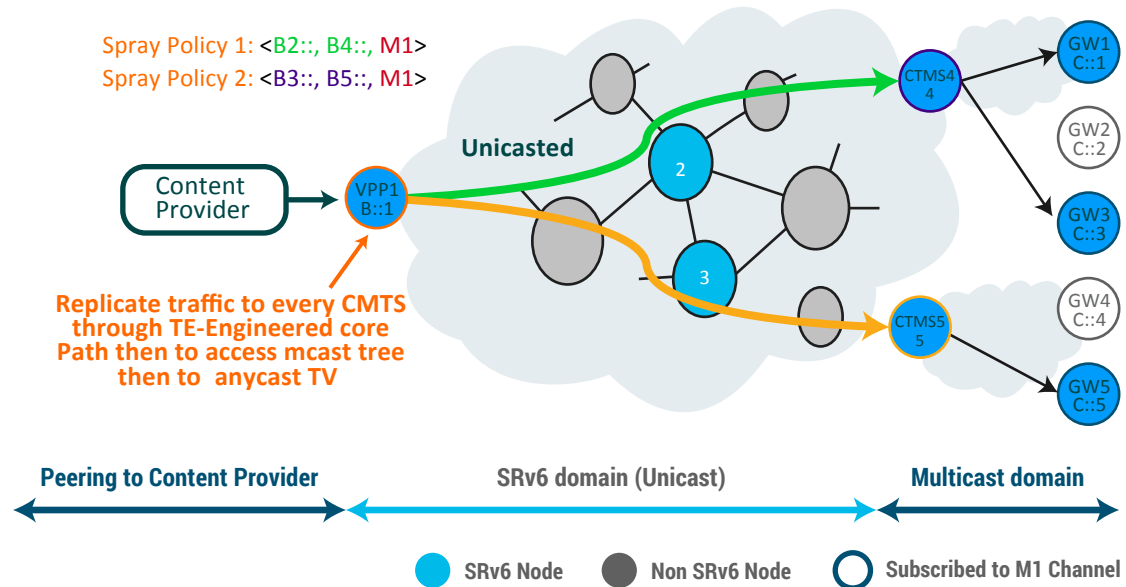
SRV6 USE CASE: MULTICAST

SRV6 enables offloading a multicast core using Unicast flows to reduce complexity. This is one of the leading service provider use cases called Spray .

In Figure 22, a service provider is peering with a content provider. The content provider replicates the information to every Cable Model Transmission System (CMTS) in different regions through a traffic-engineered core. The CMTSs will then perform the multicast function, which is done by establishing Unicast flows in the core. A spray policy is added at the headend for different flows (green and orange) regions, which will enable the flows to steer through the network as Unicast traffic. The overall complexity of the core is greatly reduced by the offload of multicast protocols in the core of the network.

---

[15] https://tools.ietf.org/html/draft-ietf-sfc-nsh-28.
[16] http://www.segment-routing.net/conferences/2017-nanog-network-as-a-computer-srv6/.

Flexible, SLA-enabled and efficient content injection without multicast core

Figure 22. SPRAY Use Case for Multicast[17]

SRV6 USE CASE: 5G

SRV6 will play an important role in 5G transport, so much so that it has the potential to replace the major tunneling protocol in the user plane called GTP-U.

In mobile networks, GTP-U is used as the tunneling protocol to carry user data in GPRS, UMTS and LTE networks (also part of 5G). Tunnels are created per session, Figure 23. The current mobile networks are rigidly fragmented between radio access, core (EPC) and service network. Tunneling techniques are used to connect these domains through anchor nodes. Such rigidness makes it difficult for the operator to optimize the data path.
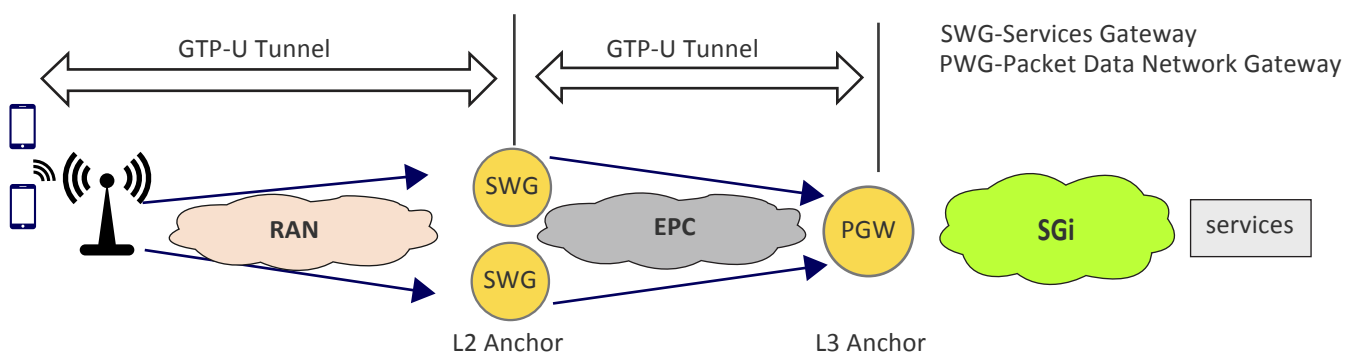

Figure 23. Current Mobile networks[18]

SRV6 can be used as a replacement for GTP-U in 5G to make the transport much simpler. TEID is used in GTP-U as an identifier to stitch different nodes. As SRV6 has SID field, so it can easily encode the TEID information therefore paving way for replacing GTP-U altogether. Not only can SRV6 replace the GTP layer, but also any underlay transport layers (for example MPLS or any other L2 tunneling protocol) paving the way for the introduction of SRV6/IPV6 as the only transport layer in 5G.

17 http://www.segment-routing.net/conferences/2017-nanog-network-as-a-computer-srv6/.

18 https://datatracker.ietf.org/meeting/99/materials/slides-99-dmm-srv6-for-mobile-user-plane/.

Network Simplification

Colllapse of multiple transport layers in one layer- IPV6, thus eliminating the need of extra tunneling protocols such as GTP

Traffic Engineering

Thanks to TE capabilities, SRV6 enables more control over data path required for diverse traffic in 5G

SRV6 5G Benefits
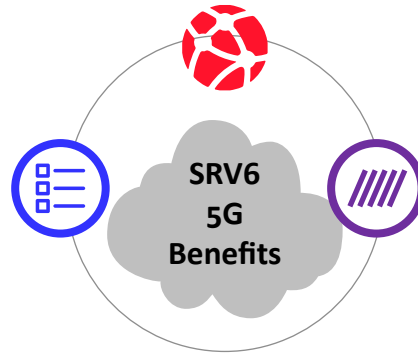
Service Chaining

SR enables service chaining and network programmability, that are extensively needed in 5G virtual core.

Network Slicing

Together with its TE and Service Chaining capabilities. SRV6 enables Network Slicing seamlessly

Figure 24. SRV6 Benefits for 5G

The 5G core is a virtualized one with complex NFV based service chain requirements in its transport layer. The inherent capability of SRV6 to provide service chaining and network programmability makes it an ideal protocol to be used in 5G environments.

Network slicing is one of the other main features of 5G transport. A common network infrastructure enables network slices depending on different SLA requirements. Each slice represents different network characteristics depending on different SLA requirements for latency, throughput and for different use cases, mobile broadband, Internet of things, etc. The benefit of creating network slicing through SRV6 is the ease through which the network slicing/virtualization can be achieved because of the SRV6 native capabilities such as tunneling, SRTE and network programmability. This eliminates the need for any additional tunneling protocols to achieve such network slicing.

USE CASE: UNIFIED FABRIC

The versatility of the use cases and features makes SR suitable to be used in any part of the network: access, metro, backbone or data center, enabling a unified fabric end to end, and eliminates the need of running different transport protocols in different part of the network.
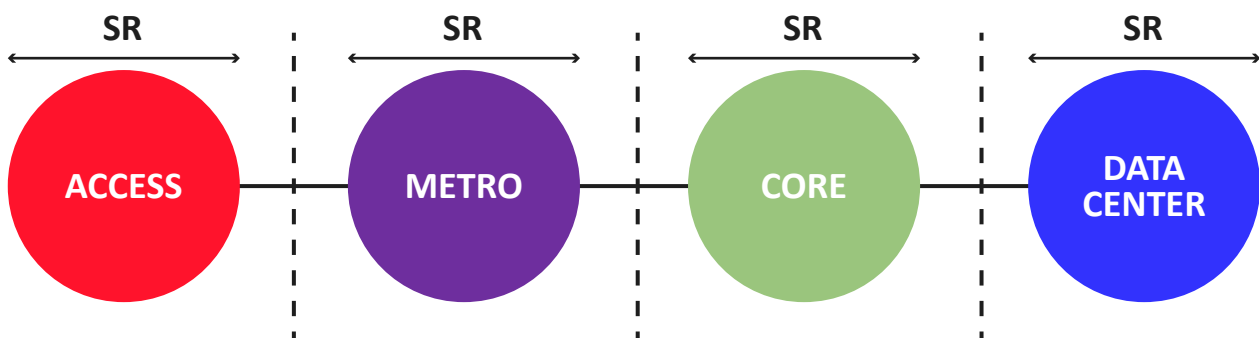
SR SR SR SR

ACCESS METRO CORE DATA CENTER

Figure 25. SR Unified Fabric

Working with a unified transport protocol such as SR in end-to-end domains has many benefits: although it makes the operation easier, the major benefit is the elimination of re-classification at the domain boundaries. For example, one common issue in a mobile backhaul network is the challenge of setting up consistent quality of service (QoS) scheme across access, metro and core.

In the absence of a consistent QoS, the network cannot fulfill consistent end-to-end SLA requirements. Issues can be avoided if there is only one transport protocol. With one unified fabric, all the advantages of SR can be utilized: end-to-end TI-LFA, end-to-end on-demand SR policy, end-to-end automatic steering and consistent 50msec recovery time no matter when the link cut happens.
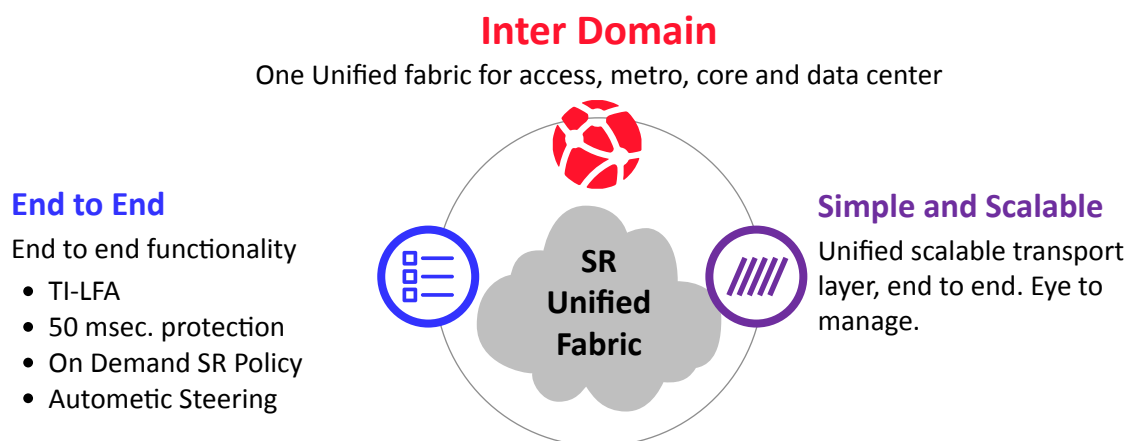
Inter Domain
One Unified fabric for access, metro, core and data center

End to End
End to end functionality
- TI-LFA
- 50 msec. protection
- On Demand SR Policy
- Autometic Steering

SR Unified Fabric

Simple and Scalable
Unified scalable transport layer, end to end. Eye to manage.

Figure 26. SR Unified Fabric Benefits

USE CASE: ENHANCED OAM FEATURES IN SEGMENT ROUTING

Segment Routing offers innovative and enhanced OAM features, including Path Monitoring System (PMS), traditional LSP Ping and traceroute tools.

In Figure 27, to monitor the path between Nodes F and D, PMS does it in two steps. In a first step it discovers all the reachable paths from F to D through the path trace message from Point F. From this information, it builds up monitoring packets that it generates from PMS with the label stack (Prefix F, Prefix D, Prefix PMS). For example, the packet travels to Node F, then Node D and returns to PMS. In this way PMS has complete visibility and status for the link between F and D. This is a novel way of path connectivity monitoring because it does not require any MPLS OAM functionality. All monitoring packets stay in the data plane; path monitoring does not require any control plane interaction in any node. Many operators prefer this way of central connectivity validation mechanism.
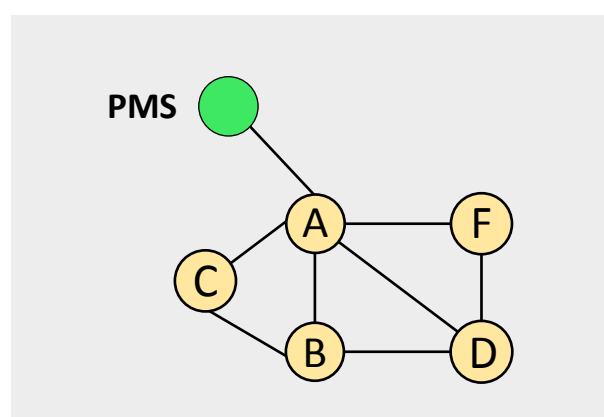

Figure 27. Path Monitoring System

Performance measurement is another important operational requirement for service providers. It is expected to meet SLAs for latency, jitter, packet loss, etc. This helps in network performance evaluation, troubleshooting and planning. RFC6374 specifies protocol mechanisms to enable the efficient and accurate measurement of these performance metrics in MPLS networks. The same methods can be applied in SR networks when used in the MPLS data plane, thus enabling the reuse of the existing performance mechanisms.

Traffic matrix collection is key to successful traffic engineering and capacity planning. Traffic collection is complex in current IP networks because it can involve many configurations on the nodes. Traffic counters are enabled (for example, NetFlow, SNMP MIB, MPLS and MIBS) and the data is collected and sent to a central engine to process and give a report on the traffic matrix. When Segment Routing is enabled in a network, the traffic collection process is automated thereby making the traffic engineering and capacity planning process much simpler and convenient.

PUBLIC USER REFERENCES

SR IS NOT JUST IN TEST LABS. IT IS ALREADY DEPLOYED AND IN PRODUCTION BY WEBSCALE COMPANIES, ENTERPRISES AND SERVICE PROVIDERS ACROSS AREAS SUCH AS WAN, METRO, DATA CENTER AND ACCESS. SOME PUBLIC REFERENCES, DRIVERS AND BENEFITS CASES ARE LISTED IN TABLE 2. NETWORK AND OPERATIONAL SIMPLICITY TOPS THE LIST OF USE CASES. THIS IS FOLLOWED BY GAINING MORE CONTROL OVER THE TRAFFIC THROUGH TRAFFIC ENGINEERING.

Company	Use Case	Drivers for Selecting Segment Routing	Benefits
Microsoft	• SWAN Project: Inter Data Center Core with SDN controller	• Different SLAs for different applications require more control over routing.	• Simplified operation. • Program only the edge instead of Core. • Removed vendor lock in.
COMCAST	• Traffic engineering in IPV6 Core • IPV6 SR Multicast	• Greater control over routing of specific applications according to their SLAs. • Running Multicast in core is complec and loads the network. The solution effectively offloads the multicast from core and instead sR unicast	• Applications engineered core. • Bandwidth savings because of offloading Multicast core • Simplicity
vodafone GERMANY	• Traffic engineering in MPLS Core	• Ability to engineer paths based on latency and application requirements	• Simplified Operation, No signaling protocols needed. • 50 percent latency reduction in paths.
Walmart	MPLS Core (Greenfield)	• Simplicity • Extensibility • Can expand cost effectively domestically and internationally.	• Simplified deployment and operation.
China unicom 中国联通	• Deployment in the backbone • Use of SDN controller with SR • (Cisco's first SR deployment in China)	• Making Network, ready for cloud. China Unicom migration to cloud only be achieved by having considtent and simple protocol across multiple domains	• Elimination of complex protocols from backbone • Centralized PCE based controller will enable China Unicm, offer cloud based services
colt	• Deployment in colt IQ network across Pan European, US and Asian packet network	• Combined SR and EVPN, to offer faster convergence, increased network availability and resiliency for any topology.	• Simplify and automate network operations and reduce operating costs
Bell	• As part of Network 3.0 transformation, Bell Canada upgraded its first four IP core routers to support SR.	• To imprrove reliability and performance of their smart core network helping better manage the overall network operations	• Increased network robustness and simplification of network operations.

Table 2. User References

CONCLUSION AND RECOMMENDATIONS

SEGMENT ROUTING IS A PROMISING TECHNOLOGY THAT CAN BE SEAMLESSLY DEPLOYED IN TODAY'S MPLS AND IPV6 NETWORKS. THE VERSATILITY OF THE TECHNOLOGY IN TERMS OF DEPLOYMENT (DISTRIBUTED VERSUS CENTRALIZED), NETWORK TYPES (DATA CENTERS OR WAN), DIVERSE USE CASES MAKES IT A GOOD CANDIDATE FOR DEPLOYMENT IN ANY KIND OF WAN, DATA CENTER, ACCESS, METRO OR VIRTUALIZED ENVIRONMENT.

THE FOLLOWING ARE RECOMMENDATIONS FOR NETWORK DESIGNERS, PLANNERS AND KEY DECISION MAKERS:

1 Assess the technological and operational pain points of current IP/MPLS networks and IPV6 networks. Recommendation: Bring on board the operation team in this exercise.

2 Understand the different use cases for SR. Every commercial deployment today has been use case driven. The biggest use case is simplicity and scalability.

3 For greenfield, it is easier and recommended to deploy SR because of the opportunities the technology offers, current IETF standards activities and success in real production networks.

4 For the brownfield environment, SR can be enabled in current IP/MPLS networks without any rip and replace strategy. It can co-exist with RSVP-TE/LDP.

5 Service providers can enable SR in their current networks on limited scale before global migration.

6 Implementing SR is a low-risk initiative considering that major protocols will be offloaded instead of burdening the network; ultimately the network will become simpler.

7 SR with a centralized controller makes sense as the core of the network is already stateless, and the controller can further take away the path computation burden off the edge nodes, enabling end-to-end control across multiple domains.

8 SRV6 enables flexible network programming. It enables the collapse of multiple layers and eliminates the need for overlay and additional protocols for service chaining, making the networks simpler to run and operate.

9 SR unified fabric leads to a simpler end-to-end transport network and reduces the number of transport protocols needed across access, metro, core and data centers.

10 SR is a way for the service providers make their networks simpler and unlock new revenue potentials.

Service providers need to reduce current complexities in their networks to compete efficiently with the webscale over-the-top providers. Network owners have only two options: either continue to grow with the complexities and lose more on capital expense and operational expense or think outside of the box with Segment Routing to solve these issues.